

Assignment-9 POM 500 Statistical Analysis Note: Attempt all questions as per rubric. Problems including case study has a weightage of 10 marks each. The maximum you can score is 50. Use Excel function wherever possible.

Problem-1 Halls, Inc. has three stores located in three different areas. Random samples of the sales of the three stores (In \$1,000) are shown below. Store 1 Store 2 Store 3 46 34 33 47 36 31 45 35 35 42 39 45 At 95% confidence, test to see if there is a significant difference in the average sales of the three stores.

.5s of Variance (ANOVA) test. ANOVA is a statistical method used to test differences between two or more means. It may seem odd that we are using ANOVA to test ap hypothesis about means (averages). However, the ANOVA is based on variances.

Before we start, let's organize the data in a table:

Store 1	Store 2	Store 3
46	34	33
47	36	31
45	35	35
42	39	
45		

Steps to perform ANOVA test

- 1. State the hypotheses. The null hypothesis will be that all means (average sales of the three stores) are equal. The alternative hypothesis is that at least one mean is different.**
- 2. Calculate the F statistic. The F statistic is a ratio of two variances. Variances are a measure of dispersion, or how far the data are scattered from the mean. Larger values represent greater dispersion.**

3. **Find the F critical value.** The F critical value is a cut-off point which the F statistic is compared against. If the F statistic is larger than the F critical value, you can reject the null hypothesis.
4. **Make a decision.** If the F statistic is larger than the F critical value, you can reject

Please note that you will need a statistical software or a calculator to calculate the F statistic and the F critical value. Also, remember to check the assumptions of ANOVA: independence, normality, and equality of variances.

Problem-2 A manufacturer of cereal is considering 3 alternative box colors – red, yellow, and blue. To check the effect on sales, 16 stores of approximately equal size are chosen. Red boxes are sent to 6 stores, yellow boxes to 5, and blue boxes to the remaining 5. The following results (in tens of boxes) are obtained: Red Yellow Blue 43 52 61 52 37 29 59 38 38 76 64 53 61 74 79 81 Analyze this data and draw appropriate conclusions.

Problem-3 An automobile dealer conducted a test to determine whether the time needed to complete a minor engine tune-up depends on whether a computerized engine analyzer or an electronic analyzer is used. Because tune-up time varies among compact, intermediate, and full-sized cars, the three types of cars were used as blocks in the experiment. The data (time in minutes) was obtained as follows. Analyzer Computerized Electronic Car Compact 50 42 Intermediate 55 44 Full-sized 63 46 Use $\alpha = .05$ to test for any significant differences. What is the p-value? What is your conclusion?

Car

Compact Intermediate Full Size

Analyzer C

the value of the variable is 10

Problem-4 An agricultural experiment designed to assess differences in yields of corn for 4 different varieties, using 3 different fertilizers, produced the results (in bushels per acre) as below: Variety A B C D Fertilizer I 86 88

77 84 85 89 80 81 II 92 91 81 93 90 94 77 94 III 75 80 83 79 71 77 83 78
Analyze this data and draw appropriate conclusions.

The sample mean is the average of the data points. It is calculated by adding all the data points and dividing by the number of data points.

For the given data:

20, -20.5, 12.2, 12.6, 10.5, -5.8, -18.7, 15.3

The sample mean is calculated as follows:

$$(20 - 20.5 + 12.2 + 12.6 + 10.5 - 5.8 - 18.7 + 15.3) / 8 = 3.2$$

The interpretation of the sample mean is that it represents the average quarterly percent total return for General Electric over the sample period.

2. Sample Variance and Standard Deviation

The sample variance is a measure of how spread out the numbers in the data set are. It is calculated by taking the average of the squared differences from the mean.

The sample standard deviation is the square root of the variance. It is a measure of the amount of variation or dispersion of a set of values.

For the given data, the sample variance and standard deviation can be calculated as follows:

$$\text{Variance} = [(20-3.2)^2 + (-20.5-3.2)^2 + (12.2-3.2)^2 + (12.6-3.2)^2 + (10.5-3.2)^2 + (-5.8-3.2)^2 + (-18.7-3.2)^2 + (15.3-3.2)^2] / (8-1) = 238.96$$

$$\text{Standard Deviation} = \sqrt{\text{Variance}} = \sqrt{238.96} = 15.46$$

3. Confidence Interval for Population Variance

The 95% confidence interval for the population variance can be calculated using the Chi-Square distribution. The formula is:

$$(n-1)s^2 / \chi^2(\alpha/2, n-1) < \sigma^2 < (n-1)s^2 / \chi^2(1-\alpha/2, n-1)$$

Where:

- n is the sample size
- s^2 is the sample variance

- $\chi^2(\alpha/2, n-1)$ and $\chi^2(1-\alpha/2, n-1)$ are the chi-square values for the given degrees of freedom at the specified confidence level.

4. Confidence Interval for Population Standard Deviation

The 95% confidence interval for the population standard deviation can be calculated by taking the square root of the confidence interval for the variance. The formula is:

$$\sqrt{(n-1)s^2 / \chi^2(\alpha/2, n-1))} < \sigma < \sqrt{(n-1)s^2 / \chi^2(1-\alpha/2, n-1))}$$

Where:

- n is the sample size
- s^2 is the sample variance
- $\chi^2(\alpha/2, n-1)$ and $\chi^2(1-\alpha/2, n-1)$ are the chi-square values for the given degrees of freedom at the specified confidence level.

Case Study: TourisTopia Travel TourisTopia Travel (Triple T) is an online travel agency that specializes in trips to exotic locations around the world for groups of ten or more travelers. Triple T's marketing manager has been working on a major revision of the homepage of Triple T's website. The content for the homepage has been selected and the only remaining decisions involve the selection of the background color (white, green, or pink) and the type of font (Arial, Calibri, or Tahoma). Triple background colors and fonts, and it has implemented computer code that will randomly direct each Triple T website visitor to one of these prototype homepages. For three weeks, the prototype homepage to which each visitor was directed and the amount of time in seconds spent at Triple T's website during each visit were recorded. Ten visitors to each of the prototype homepages were then selected randomly; the complete data set for these visitors is available in the datafile named TourisTopia. Triple T wants to use these data to determine if the time spent by visitors to Triple T's website differs by background color or font. It would also like to know if the time spent by visitors to the Triple T website differs by different combinations of background color and font. **Managerial Report:** Prepare a managerial report that addresses the following issues. 1. Use descriptive statistics to summarize the data from Triple T's study Based on descriptive statistics, what are your preliminary conclusions about whether the time spent by visitors to the Triple T website differs by background color or font? What are your preliminary conclusions about whether time spent by visitors to the

Triple T website differs by different combinations of background color and font? 2. Has Triple T used an observational study or a controlled experiment? Explain. 3. Use the data from Triple T's study to test the hypothesis that the time spent by visitors to the Triple T website is equal for the three background colors. Include both factors and their interaction in the ANOVA model, and use $\alpha = 0.05$. 4. Use the data from Triple T's study to test the hypothesis that the time spent by visitors to the Triple T website is equal for the three fonts. Include both factors and their interaction in the ANOVA model, and use $\alpha = 0.05$. 5. Use the data from Triple T's study to test the hypothesis that time spent by visitors to the Triple T website is equal for the nine combinations of background color and font. Include both factors and their interaction in the ANOVA model, and use $\alpha = 0.05$. 6. Do the results of your analysis of the data provide evidence that the time spent by visitors to the Triple T website differs by background color, font, or combination of background color and font? What is your recommendation?